Analysis of boutique arrays: A universal method for the selection of the optimal data normalization procedure

BARBARA USZCZYŃSKA[1], JOANNA ZYPRYCH-WALCZAK[2], LUIZA HANDSCHUH[1,3], ALICJA SZABELSKA[2,5], MACIEJ KAŹMIERCZAK[3], WIESŁAWA WORONOWICZ[1], PIOTR KOZŁOWSKI[1], MICHAŁ M. SIKORSKI[1], MIECZYSŁAW KOMARNICKI[3], IDZI SIATKOWSKI[2] and MAREK FIGLEROWICZ[1,4]

[1]Institute of Bioorganic Chemistry, Polish Academy of Sciences, 61-704 Poznań; [2]Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, 60-637 Poznań; [3]Department of Hematology, Poznań University of Medical Sciences, 60-569 Poznań; [4]Institute of Computing Science, Poznań University of Technology, 61-138 Poznań, Poland; [5]Functional Genomics Center Zurich, Swiss Federal Institute of Technology (ETH Zurich) and the University of Zurich, CH-8057 Zurich, Switzerland

Received March 21, 2013; Accepted May 28, 2013

DOI: 10.3892/ijmm.2013.1443

Abstract. DNA microarrays, which are among the most popular genomic tools, are widely applied in biology and medicine. Boutique arrays, which are small, spotted, dedicated microarrays, constitute an inexpensive alternative to whole-genome screening methods. The data extracted from each microarray-based experiment must be transformed and processed prior to further analysis to eliminate any technical bias. The normalization of the data is the most crucial step of microarray data pre-processing and this process must be carefully considered as it has a profound effect on the results of the analysis. Several normalization algorithms have been developed and implemented in data analysis software packages. However, most of these methods were designed for whole-genome analysis. In this study, we tested 13 normalization strategies (ten for double-channel data and three for single-channel data) available on R Bioconductor and compared their effectiveness in the normalization of four boutique array datasets. The results revealed that boutique arrays can be successfully normalized using standard methods, but not every method is suitable for each dataset. We also suggest a universal seven-step workflow that can be applied for the selection of the optimal normalization procedure for any boutique array dataset. The described workflow enables the evaluation of the investigated normalization methods based on the bias and variance values for the control probes, a differential expression analysis and a receiver operating characteristic curve analysis. The analysis of each component results in a separate ranking of the normalization methods. A combination of the ranks obtained from all the normalization procedures facilitates the selection of the most appropriate normalization method for the studied dataset and determines which methods can be used interchangeably.

Correspondence to: Dr Luiza Handschuh or Dr Piotr Kozłowski, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznań, Poland
E-mail: luizahan@ibch.poznan.pl
E-mail: kozlowp@yahoo.com

Abbreviations: AML, acute myeloid leukemia; AUC, area under the curve; DEA, differential expression analysis; FDR, false discovery rate; PCR, polymerase chain reaction; ROC curve, receiver operating characteristic curve

Key words: DNA microarrays, boutique arrays, normalization, data analysis, R Bioconductor, acute myeloid leukemia

Introduction

Despite the dynamic development of deep sequencing technologies, microarrays are still commonly used in genomic research (1-5). Currently, DNA microarrays are mainly used for genotyping (6-9), gene expression profiling (10-12) and microRNA screening (13-15). In medicine, microarrays are used to determine the complexity and heterogeneity of diseases, to facilitate disease classification and to predict therapeutic outcomes (8,16-22).

Microarrays provide a large amount of useful information, but are accompanied by inherent noise and systematic errors (23-26). No microarray experiment is free from variation introduced during sample preparation, hybridization, washing and scanning (24,27,28). Spotted arrays are burdened with technical defects that occur during their printing; these defects manifest as differences in spot size and shape and/or shifts of spots, rows or whole print-tips (24,28). In two-color assays, additional bias is introduced by uneven dye incorporation and by differences in the signal dynamic range and the sensitivity of dyes to photobleaching (23,24,29). Therefore, the major challenge in microarray analysis is data pre-processing, which includes normalization and background correction (24,31-36). Background correction enables the removal of ambient, non-specific signals and spatial background heterogeneity across the array (34). Normalization eliminates any non-biological variation in the foreground signals and calibrates the distributions and values of the signal intensities within and between arrays (24,32,37,38). Currently, several normalization methods

are available to address different types of bias and offer various solutions. The global normalization methods used in transcriptome profiling are based on the assumption that differentially expressed genes constitute a small fraction of the genes that are represented on the array (38,39). In addition, a balance between the upregulated and downregulated genes is expected (13,30,40). These assumptions are usually fulfilled in whole-genome analysis, although asymmetric gene expression often occurs in cancer cells or in genetically modified models (41). Simple global methods, such as median or mean normalization, assume that all the genes represented on an array are equally affected by the bias (39). Since this assumption is usually incorrect, a more highly recommended method for global and print-tip-based microarray normalization is local regression (32,42,43). This alternative to global normalization is an approach that favors selected sets of control probes that are expected to generate uniform signal intensities within and between arrays, e.g., spike-in probes (13,41,44-47).

Due to the profound effect of normalization on the results of microarray data analysis, e.g., the selection of differentially expressed genes (27,48,49), the normalization strategy must be carefully considered. Even in the case of Affymetrix chips, which are known for their uniformity and reproducibility, the normalization significantly influences the results of the data analysis. Previous studies have shown that the lists of differentially expressed genes obtained as a result of the application of various normalization algorithms to Affymetrix data only partially overlap (44,46,50). This data pre-processing issue is much more complex in the case of custom- and home-made microarrays, which often present atypical and unique features. Although a wide spectrum of ready-to-use arrays is offered in the market, there is still a need to produce arrays with specific designs that are devoted to studies of less-known organisms and/or dedicated to a specific biological or biomedical issue (51-55). Such microarrays are often termed 'boutique arrays' (13,47,56). These arrays are usually spotted using the contact printing technique and designed for two-color hybridization (57,58).

In this study, we present a universal seven-step workflow for the selection of the optimal normalization procedure for the analysis of a specific boutique array dataset. Using statistical functions implemented in R (http://www.R-project.org) and Bioconductor (30,59,60), we used ten methods for double-channel normalization and three methods for single-channel normalization for the pre-processing of three datasets from our laboratory [the acute myeloid leukemia (AML), Allergy and Asthma datasets] and the dataset described in the study by Oshlack *et al* (47).

Materials and methods

Microarray data sources. Four sets of small spotted microarray datasets were used to compare the normalization methods. These arrays include three sets of our own data (the AML dataset, the Allergy and Asthma dataset) and the data published in the study by Oshlack *et al* (47). A summary of the experiments and data structures is presented in Table I.

Normalization. All the computations and analyses were performed using R 2.13 and Bioconductor 2.7. The normalization of the raw data (.gpr files) was performed using the limma (61), snm (62), vsn (63), nnNorm (64), optimized local intensity-dependent normalization (OLIN) (65), marray (30) and TurboNorm (66) packages (downloaded from http://www.bioconductor.org/packages/release/bioc/) according to the instructions enclosed in the manuals. The limma, vsn and snm packages were also used for the single-channel normalization of double-channel data. A summary of the normalization methods is shown in Table II. The diagnostic plots (MAplots and box plots) used in the analysis were generated using the limma and graphics package.

Comparison of the normalization methods

Bias and variance. To evaluate the normalization methods that are used for the normalization of double-channel microarray data, we used the formal approach proposed in the study by Argyropoulos *et al* (67). These authors determined the bias and variance values for spike-in control spots using the following formulae:

$$\text{bias}_i = \sqrt{\sum_j \sum_k \left(\log_2\left(R_{i,j,k}/G_{i,j,k}\right)\right)^2 / n} \quad [1]$$

and

$$\text{variance}_i = \sum_j \sum_k \left(\log_2\left(R_{i,j,k}/G_{i,j,k}\right) - \left\langle \log_2\left(R_{i,j,k}/G_{i,j,k}\right)\right\rangle_i\right)^2 /(n-1) \quad [2]$$

where $R_{i,j,k}$ and $G_{i,j,k}$ are, respectively, the red and green intensity values, of a spot in the k-th replicate of the i-th control probe from the j-th microarray, n = j*k and $\left\langle \log_2\left(R_{i,j,k}/G_{i,j,k}\right)\right\rangle_i$ is the mean value of the red and green intensity log-ratio of the i-th control probe.

We applied a similar strategy to compare the single-channel normalization methods. Theoretically, the intensity of each control spot should be the same regardless of the fluorescent dye used for the labeling, i.e.,

$$R_{i,j,k}/G_{i,j,k} \to 1 \Rightarrow \log_2\left(\frac{R_{i,j,k}}{G_{i,j,k}}\right) \to 0. \quad [3]$$

This assumption also implies that the red and green channel intensities of each control probe should be constant. Therefore $R_{i,j,k} \to$ const and $G_{i,j,k} \to$ const for each k-th replicate of the i-th control probe in the j-th microarray. As we were interested only in the red channel intensities, we concluded that, $R_{i,j,k} \to \bar{R}_i$, where $\bar{R}_i$ is the mean intensity of the red channel of the i-th control probe. Taking the logarithm of the above expression, we obtained the following formula:

$$\log_2 R_{i,j,k} - \log_2 \bar{R}_i \to 0 \Rightarrow \log_2\left(\frac{R_{i,j,k}}{\bar{R}_i}\right) \to 0. \quad [4]$$

Thus, to calculate the bias in the single-channel normalization in equations [1] and [2], we used the expression from equation [4] rather than the $\log_2(R_{i,j,k}/G_{i,j,k})$ expression. Thus, for the single-channel normalization, we obtained the following formulae:

$$\text{bias}_i = \sqrt{\sum_j \sum_k \left(\log_2\left(R_{i,j,k}/\bar{R}_i\right)^2\right)/n} \quad [5]$$

and

$$\text{variance}_i = \sum_j \sum_k \left(\log_2\left(R_{i,j,k}/\bar{R}_i\right) - \left\langle \log_2\left(R_{i,j,k}/\bar{R}_i\right)\right\rangle_i\right)^2 /(n-1). \quad [6]$$

The same operations can be repeated for the analysis of the green channel.

Table I. Summary of the microarray experiments and dataset structures.

Features	AML dataset	Allergy dataset	Asthma dataset	Oshlack dataset
No. of arrays	40	14	14	6
No. of unique probes	919	208		146
Type of probes	Oligo ~50 nt, 5' amino-modified (Ocimum Biosolutions)	Oligo ~50 nt, 5' amino-modified (Ocimum Biosolutions)		PCR fragments corresponding to cDNAs
No. of spike-in probes and spike RNAs	8 (ArrayControl, Ambion)	8 (ArrayControl, Ambion)		MSP probes (pooled clones of cDNA library)
No. of replicate spots	3	6		2
No. of spots per print-tip (and No. of print-tips)	81 (in 40 print-tips)	144 (in 8 print-tips) or 48 (in 4 last print-tips)		462 (in 24 print-tips)
Design	Common reference (green channel) vs. studied samples or control samples (red channel)	Control 1 vs. Studied 1; Control 2 vs. Studied 2; … Control n vs. Studied n;		Pairwise hybridizations (mice knock-out cells vs. control cells)
Dye pairs	Alexa 555/Alexa 647	Alexa 555/Alexa 647		Cy3/Cy5
Array design ID[a]	A-MEXP-2220	A-MEXP-2209		[b]
Experiment ID[a]	E-MEXP-3647	E-MEXP-3633	E-MEXP-3646	[b]

[a]The accession numbers for the AML, Allergy and Asthma datasets were obtained from the ArrayExpress database (http://www.ebi.ac.uk/arrayexpress). [b]The Oshlack raw data can be found at http://bioinf.wehi.edu.au/folders/boutique/. AML, acute myeloid leukemia; PCR, polymerase chain reaction.

Differential analysis. In the case of the AML dataset, a differential analysis was performed after each normalization procedure. A t-test was used to identify the differentially expressed genes between the AML and the control samples. The final P-values were calculated using the false discovery rate (FDR) correction, based on the procedure introduced in the study by Benjamini and Hochberg (68).

Receiver operating characteristic (ROC) curves and area under the curve (AUC) values. ROC curves were applied to determine the effectiveness of the sample classification based on the gene signature classifiers identified in the AML dataset after each normalization procedure. Stacking regression was used to improve the prediction accuracy (69). To estimate the power of the classifiers, the AUC values were determined. All the functions necessary for the ROC curve tracing and AUC calculations were obtained from the pROC package (70).

Results

Our laboratory participates in different projects that are based on microarray analysis. Some of these projects demand the design, production and analysis of small, atypical arrays that are devoted to specific biological issues. As is true for all boutique arrays, these arrays require a more individualized approach in their data processing to minimize technical variation without losing biologically relevant information. Thus, we wished to develop a more complex and general approach to solve the issue of data normalization and to support the choice of the optimal normalization method for any dataset obtained from the analysis of small spotted microarrays. To achieve this aim, we normalized three datasets that were generated in our laboratory (the AML, Allergy and the Asthma datasets) and compared these with the data from the study published by Oshlack *et al* (47), which were collected from boutique B-cell microarrays (Table I). We compared the effectiveness of 13 microarray data normalization methods and developed a seven-step workflow that can be applied for the selection of the optimal normalization method for any dataset.

Normalization. In the case of two-color data, we focused on within-array normalization. Due to the high proportion of background in the foreground intensities (Fig. 1), a simple background correction (subtract method, limma package) was performed. To minimize the effects of other pre-processing steps, we did not perform between-array normalization.

First, we focused on the most common class of normalization methods, which are based on robust locally weighted regression. These intensity-dependent methods, which are usually termed 'loess' or 'lowess', have become a standard approach in microarray data pre-processing due to their efficiency and flexibility. We selected seven loess methods implemented in four packages in the Bioconductor software: limma, marray, OLIN and TurboNorm. The limma package offers global loess normalization (hereafter termed 'Loess'), print-tip loess (hereafter termed 'Ploess') and spike-based loess (hereafter termed 'Spike'). The loess method from the marray package (hereafter termed 'LoessM') is equivalent to the Loess method in the limma package. One of the normalization algorithms provided by the OLIN package is OLIN (optimized local intensity-dependent normalization) that we used in two different versions. One, which is hereafter termed 'Olin_c', takes into consideration the X and Y coordinates of the spots, whereas the other, which is

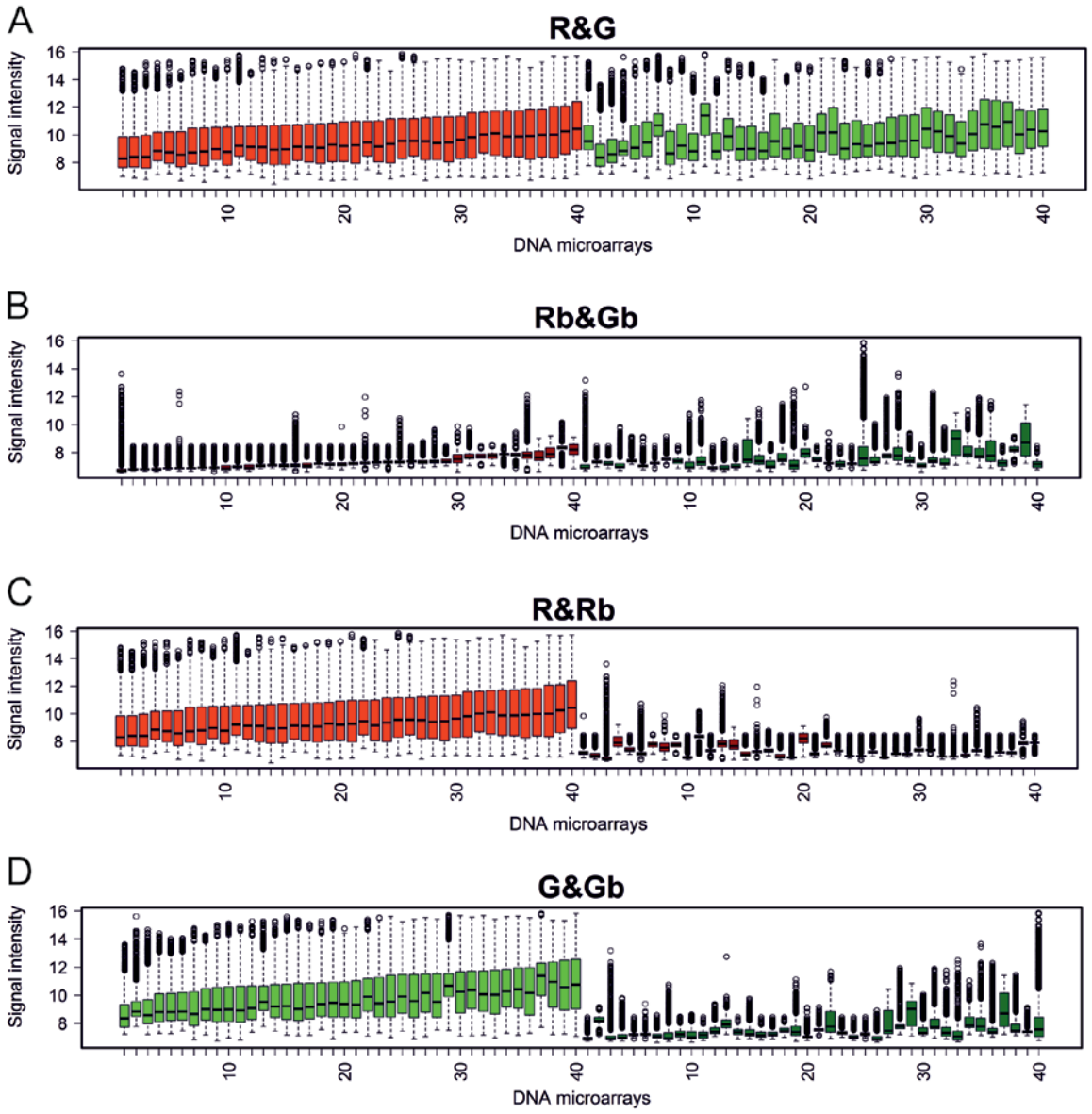

Figure 1. A set of box plots showing the raw foreground and background intensities in the red and green channels for 40 microarrays from the acute myeloid leukemia (AML) dataset. (A) Foreground intensities for both channels arranged according to increasing overall intensity within the red channel. (B) Background intensities for both channels arranged according to increasing overall intensity within the red channel. (C) Foreground and background intensities for the red channel arranged according to increasing overall red foreground intensity. (D) Foreground and background intensities for the green channel arranged according to increasing overall green foreground intensity. Colors: red, red channel intensities; green, green channel intensities. G, green; R, red; Rb, red background; Gb, green background.

hereafter termed 'Olin', does not require this information. The TurboNorm package, which is one of the newest packages, offers a loess-based normalization (hereafter termed 'Turbo') that is faster than the current loess algorithms.

Three other normalization approaches that we tested are included in the vsn, nnNorm and snm packages in Bioconductor. The method in the vsn package (hereafter termed 'Vsn2') is based on the assumption of a constant coefficient of variation. The supervised normalization method presented in the snm package (hereafter termed 'Snm2') uses all available information about the experiment to remove the bias introduced by different variables. A completely novel microarray normalization strategy, which is implemented in the nnNorm package (hereafter termed 'Nn'), uses a neural network algorithm to correct the intensity and spatial bias in the microarray data.

The design of a microarray experiment occasionally enables the conversion of double-channel to single-channel data. An example is the AML dataset, in which the common reference was used only in the green channel. To perform the single-channel data normalization, we used three methods from the limma (quantile method, hereafter termed Q), vsn (hereafter termed Vsn1) and snm (hereafter termed Snm1) packages.

Altogether, we tested ten double-channel and three single-channel normalization methods from seven Bioconductor packages (Table II).

In practice, the choice of a normalization method often relies on visualizations, i.e., the comparison of diagnostic plots, such as box plots or scatter plots, before and after the normalization. A useful type of scatter plot is the MAplot, which traces the R/G log-ratio (M-values) against the overall

Table II. Summary of applied normalization methods.

Name	Name of function	Bioconductor package	Short description of method	Additional features	Authors/(Refs.)
Q	NormalizeBetweenArrays	limma	Normalization based on sorting and transformation of foreground intensities to obtain the same signal intensity distribution on each array.	Usually applied to normalization of two-color data. Also enables single-channel normalization.	Bolstad *et al* (31) Smyth and Speed (32)
Loess	NormalizeWithinArrays	limma	Normalization model based on robust locally weighted regression.	-	Smyth and Speed (32)
Spike	NormalizeWithinArrays	limma	Normalization model based on robust locally weighted regression.	Global loess normalization based on spike-in controls.	Smyth and Speed (32)
Ploess	NormalizeWithinArrays	limma	Normalization model based on robust locally weighted regression.	Global loess normalization applied separately to each print-tip.	Smyth and Speed (32)
LoessM	maNorm	marray	Normalization model based on robust locally weighted regression.	Offers a flexible approach for the location and scale normalization of M-values.	Yang *et al* (72)
Vsn2[a] Vsn1[a]	vsn2	vsn	Model based on the assumption of a constant coefficient of variation.	Also enables single-channel normalization of two-color data.	Huber *et al* (63)
Nn	maNormNN	nnNorm	Normalization based on neural networks models using average log-intensity (A) values and pseudo spatial spot coordinates (X and Y) as predictors.	Resistant to outliers.	Tarca *et al* (64)
Olin	olin	olin	Includes two normalization schemes based on iterative local regression and model selection.	Does not require X and Y spot coordinates.	Futschik and Crompton (65)
Olin_c	olin	olin	Includes two normalization schemes based on iterative local regression and model selection.	Requires X and Y spot coordinates.	Futschik and Crompton (65)
Turbo	pspline	TurboNorm	Normalization based on a weighted P-spline scatter plot smoother.	Simpler and faster than current loess algorithms.	van Iterson *et al* (66)
Snm2[a] Snm1[a]	snm	snm	Supervised normalization method that defines and models different sources of variation based on the study design.	Also enables single-channel normalization of two-color data.	Mecham *et al* (62)

[a]The 1 and 2 suffixes indicate single and double-channel normalization, respectively. Q, quantile.

intensity of each spot (A-values). Examples of MAplots, which were generated for one microarray from the AML dataset, are shown in Fig. 2. An MAplot of raw microarray data usually takes a typically curved shape (Fig. 2A) that is straightened as

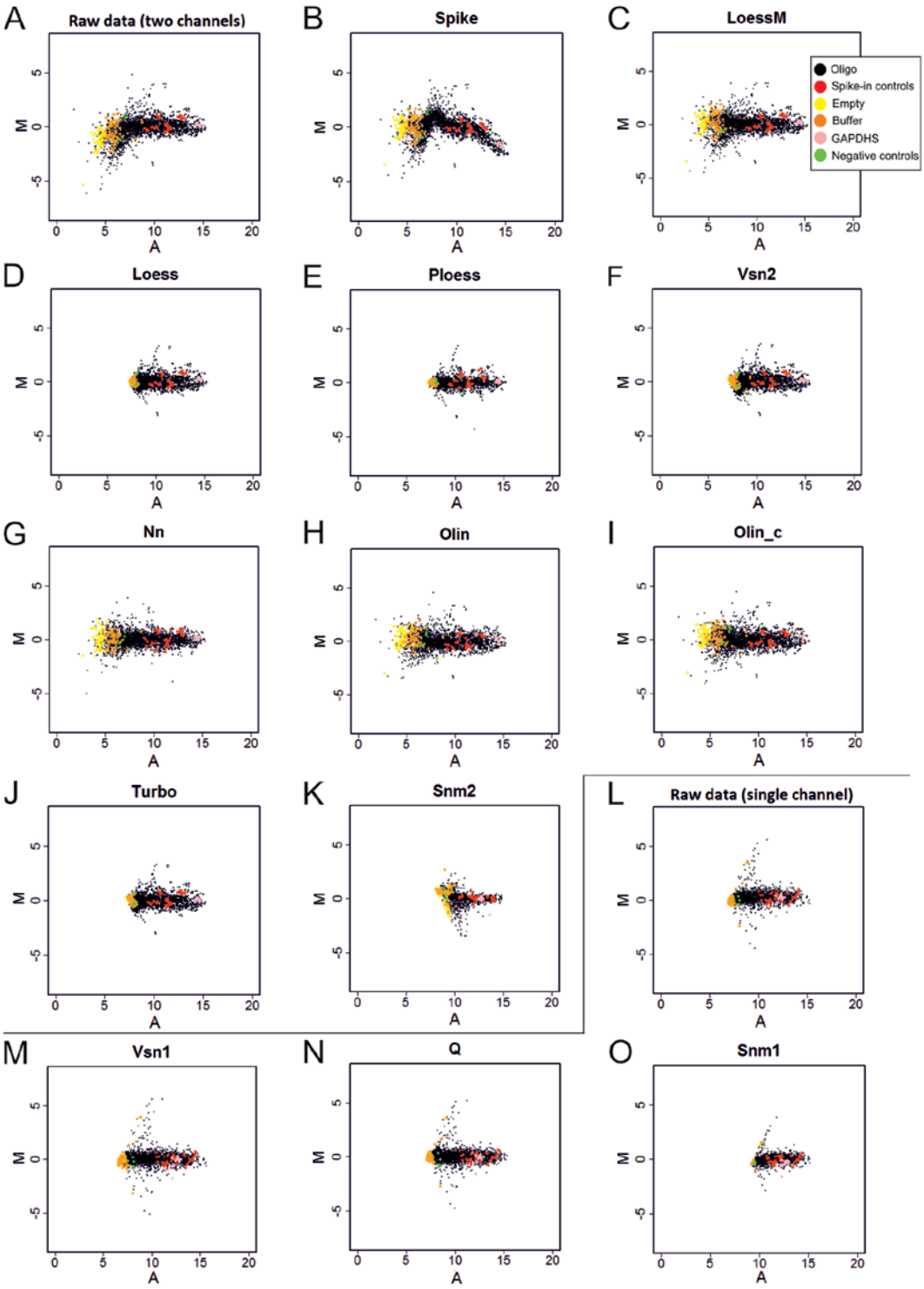

Figure 2. MAplots obtained using different normalization methods for the normalization of one microarray from the acute myeloid leukemia (AML) dataset. The MAplots that indicate the double-channel normalization methods: (A) Double-channel raw data; (B) Spike (global loess normalization method based on spike-in controls, limma package); (C) LoessM (global loess normalization method, marray package); (D) Loess (global loess normalization method, limma package); (E) Ploess (global loess normalization method applied separately to each print-tip, limma package); (F) Vsn2 (global normalization method, vsn package); (G) Nn (normalization method based on neural networks models, nnNorm package); (H) Olin (global loess normalization method, OLIN package); (I) Olin_c (global loess normalization method including X and Y spot coordinates, Olin package); (J) Turbo (global loess normalization method, TurboNorm package); and (K) Snm2 (supervised normalization method, snm package). The MAplots that indicate the single-channel normalization methods: (L) Single-channel raw data; (M) Vsn1 (global normalization method, vsn package); (N) Q (quantile global normalization method, limma package); and (O) Snm1 (supervised normalization method, snm package). The control probes are indicated by the colors described in the legend.

a result of efficient normalization of the M-values (Fig. 2C-J). It appears that all of the tested methods, with the exception of three (the Spike, Snm2 and Snm1 methods, which are shown in Fig. 2B, K and O, respectively), were able to normalize the AML raw data. However, it is challenging to decide which of the methods is optimal. The interpretation of the plots is

Table III. Rank of the bias values obtained using the 13 normalization methods for the normalization of the four tested datasets.

Normalization method	AML dataset Rank (bias values)	Allergy dataset Rank (bias values)	Asthma dataset Rank (bias values)	Oshlack dataset[a] Rank (bias values)
Double-channel normalization				
Spike	10 (4.76)	9 (0.83)	8 (0.87)	10 (1.52)
LoessM	8 (0.76)	7 (0.76)	9 (0.93)	5 (0.54)
Loess	2 (0.56)	1 (0.40)	3 (0.65)	2 (0.21)
Ploess	7 (0.73)	4 (0.51)	1 (0.54)	1 (0.20)
Vsn2	3 (0.67)	3 (0.48)	5 (0.80)	4 (0.37)
Nn	6 (0.69)	8 (0.80)	7 (0.85)	8 (0.69)
Olin	3 (0.67)	5 (0.69)	4 (0.73)	7 (0.62)
Olin_c	5 (0.68)	6 (0.72)	6 (0.83)	6 (0.60)
Turbo	1 (0.55)	2 (0.42)	2 (0.63)	3 (0.26)
Snm2	9 (1.22)	10 (1.37)	10 (1.27)	9 (0.80)
Single-channel normalization				
Vsn1	2 (0.79)	1 (0.55)	2 (0.55)	3 (0.58)
Q	1 (0.74)	2 (0.56)	1 (0.54)	1 (0.23)
Snm1	3 (1.03)	3 (1.13)	3 (1.20)	2 (0.39)

AML, acute myeloid leukemia; Q, quantile. [a]Dataset descrbibed in the study by Oshlack *et al* (47).

intuitive and depends on the experience of the researcher. Moreover, the analysis of many samples and several normalization methods would generate a large amount of diagnostic plots. The processing of the AML dataset, which contains 40 microarrays and was normalized using 13 methods, generated 520 MAplots. Thus, it is very difficult to choose the appropriate normalization method based only on these plots. We therefore suggest the application of other criteria for the comparison of the normalization procedures and the treatment of the plots as secondary determinants.

Bias and variance calculation. Most arrays contain a set of 'invariant' probes, e.g., housekeeping genes or control probes for external RNA. The calculation of the bias and variance of the expression values for these probes, which is described in detail in Materials and methods (equations [1], [2], [5] and [6]), appears to be the most obvious measure to differentiate among normalization methods. Following the procedure described in the study by Argyropoulos *et al* (67), we also assumed that the normalization method that gives the lowest bias and variance values for the control spots should be selected as the most appropriate method. Using equations [1] and [2] for double-channel normalization and equations [5] and [6] for single-channel normalization, we calculated the bias and variance values for the control probes present in each of the four datasets. The results are shown in Tables III and IV. The ranking in both tables was determined using the mean of the bias and variance values calculated for all of the control probes, i.e., the method with the lowest bias or variance was ranked as 1.

Based on the data presented in Tables III and IV, we were able to conclude that the best results for the double-channel normalizations were obtained using methods that are based on locally weighted regression: Turbo, Loess and Ploess. In the case of the dataset from the study by Oshlack *et al* (47) and the Asthma dataset, the best result was achieved using the Ploess method; however, the Loess and Turbo methods exhibited only slightly worse bias and variance values. The methods with the highest bias and variance were the Spike and Snm2 methods, which also produced the worst MAplots.

The analysis of the single-channel normalization methods revealed that the Q method exhibited the best performance in the normalization of the dataset from the study by Oshlack *et al (*47), as well as the AML and Asthma datasets. In the case of our three datasets (AML, Allergy and Asthma), the differences between the Q and Vsn1 methods are negligible. The worst method for the normalization of our three datasets was Snm1, whereas the worst method for the normalization of the dataset from the study by Oshlack *et al* (47) was Vsn1. However, the differences between the worst and the best methods were not as large as the differences obtained with the double-channel normalization methods.

Differential expression analysis (DEA). Another approach that can be used to select the optimal normalization method takes into account the genes that are identified as differentially expressed in the compared experimental conditions. We performed a differential analysis of the AML dataset, which contained samples that were obtained from AML patients and healthy volunteers. To determine the differences between these two groups, we used a t-test with the FDR correction for multiple testing. We selected genes that exhibited adjusted P-values <0.05. Each normalization method resulted in different numbers of genes that were identified as differentially expressed, as presented in Table V. The majority of the methods, with the exception of the Snm2, Spike and Snm1 methods, enabled the selection of approximately 200 genes, which constitutes approximately 20% of the genes present on the array. This number was logical as we expected that at

Table IV. Rank of the variance values obtained using the 13 normalization methods for the normalization of the four tested datasets.

Normalization method	AML dataset Rank (variance values)	Allergy dataset Rank (variance values)	Asthma dataset Rank (variance values)	Oshlack dataset[a] Rank (variance values)
Double-channel normalization				
Spike	10 (46.41)	9 (1.49)	9 (1.08)	10 (4.60)
LoessM	7 (0.59)	7 (0.73)	8 (0.81)	5 (0.43)
Loess	2 (0.29)	1 (0.15)	3 (0.41)	2 (0.08)
Ploess	9 (2.38)	4 (0.25)	1 (0.29)	1 (0.07)
Vsn2	3 (0.39)	3 (0.18)	5 (0.56)	4 (0.14)
Nn	6 (0.52)	8 (0.84)	7 (0.71)	8 (0.57)
Olin	4 (0.41)	5 (0.60)	4 (0.56)	7 (0.46)
Olin_c	5 (0.46)	6 (0.66)	6 (0.64)	6 (0.45)
Turbo	1 (0.28)	2 (0.16)	2 (0.39)	3 (0.09)
Snm2	8 (1.54)	10 (2.04)	10 (1.55)	9 (0.69)
Single-channel normalization				
Vsn1	2 (0.63)	1 (0.34)	2 (0.33)	3 (0.32)
Q	1 (0.56)	2 (0.35)	1 (0.34)	1 (0.06)
Snm1	3 (1.07)	3 (1.26)	3 (1.46)	2 (0.18)

AML, acute myeloid leukemia; Q, quantile. [a]Dataset descrbibed in the study by Oshlack *et al* (47).

least this number of genes would be deregulated as a result of disease. The worst performance of the Spike method is possibly due to its inability to efficiently remove technical variations. An extremely low number of differentially expressed genes (only 35) was identified in the AML dataset after the application of the double-channel normalization method included in the snm package (Snm2).

To reasonably compare the results of differential analysis, we analyzed the contents of the lists. First, we determined how many and which genes were shared between the lists that were generated after the application of the different normalization methods. Table V contains the number of common genes for each pair of methods, whereas Fig. 3 presents how the number of common genes decreases after the addition of successive methods, which are arranged according to the level of similarity.

As shown in Table V, the Snm2 method is the most divergent. Only one to five genes (up to 14%) from the set identified after the application of the Snm2 normalization method were also included in the other sets. Of the 87 genes selected as differentially expressed following Spike normalization, which was the second most divergent method, up to 32% of the genes were shared with the other lists. The highest similarity (183 shared genes) was observed between the two global loess methods from the limma and marray packages: Loess and LoessM. The following methods more similar to each other were other normalization methods based on local regression: Turbo, Olin_c and Olin. Eight double-channel methods (all of the tested methods apart from Snm2 and Spike) shared 59 genes (Fig. 3). The analysis of the single-channel methods revealed that the Q and Vsn1 normalizations gave very similar outcomes (89-92% common genes). However, we did not find a single gene that was found in all of the 13 lists (Fig. 3).

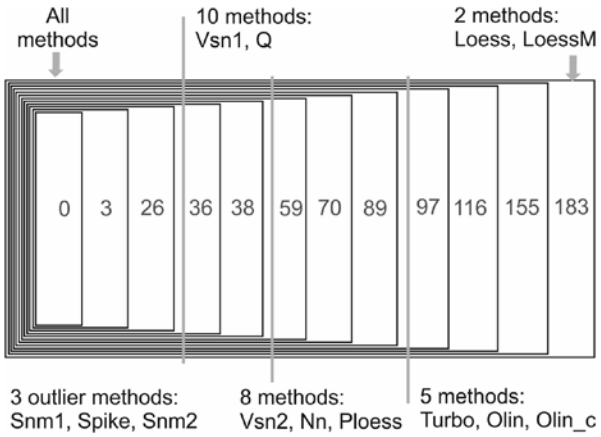

Figure 3. The numbers of differentially expressed genes shared between the lists generated after the normalization of the acute myeloid leukemia (AML) dataset using each tested normalization method. The two most concordant methods are Loess and LoessM (183 shared genes). The following methods were added according to the similarity of their gene lists: Turbo (155 genes shared with Loess and LoessM), Olin_c (116 genes shared with Loess, LoessM and Turbo), Olin (97 genes shared with the previously added methods), Vsn2 (89 genes shared with the previously added methods), Nn (70 genes shared with the previously added methods), Ploess (59 genes shared with the previously added methods), Vsn1 (38 genes shared with the previously added methods), quantile (Q; 36 genes shared with the previously added methods), Snm1 (26 genes shared with the previously added methods), Spike (three genes shared with the previously added methods) and Snm2 (0 genes shared with the previously added methods). The lines indicate the most critical border points.

To investigate the sensitivity and specificity of the normalization methods, we selected the subsets of microarray probes that could be treated as positive and negative controls. For the set of positive controls, we selected genes that were validated

Table V. Pairwise comparison of the genes in the AML dataset identified as differentially expressed using the different normalization methods.

	Double-channel methods, No. of genes (%)										Single-channel methods, No. of genes (%)		
	Spike	LoessM	Loess	Ploess	Vsn2	Nn	Olin	Olin_c	Turbo	Snm2	Vsn1	Q	Snm1
Spike	**87**	28	21	18	24	24	21	25	23	1	25	25	16
	(100)	(32)	(24)	(21)	(28)	(28)	(24)	(29)	(26)	(1)	(29)	(29)	(18)
LoessM	28	**239**	183	122	147	136	153	165	173	1	95	94	82
	(12)	**(100)**	(77)	(51)	(62)	(57)	(64)	(69)	(72)	(0.4)	(40)	(39)	(34)
Loess	21	183	**202**	114	141	116	121	130	165	0	76	77	70
	(10)	(91)	**(100)**	(56)	(70)	(57)	(60)	(64)	(82)	(0)	(38)	(38)	(35)
Ploess	18	122	114	**181**	106	115	111	113	117	3	76	74	54
	(10)	(67)	(63)	**(100)**	(59)	(64)	(61)	(62)	(65)	(2)	(42)	(41)	(30)
Vsn2	24	147	141	106	**166**	108	116	127	148	0	93	91	76
	(14)	(89)	(85)	(64)	**(100)**	(65)	(70)	(77)	(89)	(0)	(56)	(55)	(46)
Nn	24	136	116	115	108	**242**	128	132	119	5	85	87	58
	(10)	(56)	(48)	(48)	(45)	**(100)**	(53)	(55)	(49)	(2)	(35)	(36)	(24)
Olin	21	153	121	111	116	128	**211**	168	119	4	92	91	64
	(10)	(73)	(57)	(53)	(55)	(61)	**(100)**	(80)	(56)	(2)	(44)	(43)	(30)
Olin_c	25	165	130	113	127	132	168	**209**	131	4	91	92	71
	(12)	(79)	(62)	(54)	(61)	(63)	(80)	**(100)**	(63)	(2)	(44)	(44)	(34)
Turbo	23	173	165	117	148	119	119	131	**204**	0	86	87	75
	(11)	(85)	(81)	(57)	(72)	(58)	(58)	(64)	**(100)**	(0)	(42)	(43)	(37)
Snm2	1	1	0	3	0	5	4	4	0	**35**	4	6	2
	(3)	(3)	(0)	(9)	(0)	(14)	(11)	(11)	(0)	**(100)**	(11)	(17)	(6)
Vsn1	25	95	76	76	93	85	92	91	86	4	**196**	180	78
	(13)	(48)	(39)	(39)	(47)	(43)	(47)	(46)	(44)	(2)	**(100)**	(92)	(40)
Q	25	94	77	74	91	87	91	92	87	6	180	**203**	79
	(12)	(46)	(38)	(36)	(45)	(43)	(45)	(45)	(43)	(3)	(89)	**(100)**	(39)
Snm1	16	82	70	54	76	58	64	71	75	2	78	79	**118**
	(14)	(69)	(59)	(46)	(64)	(49)	(54)	(60)	(64)	(2)	(66)	(67)	**(100)**

The number of genes identified after each normalization method is indicated in the diagonal (bold print). The data above and below the diagonal indicate the number of genes shared by the two respective methods. The percentage of shared genes (numbers in parentheses) is relative to the total number of genes identified using the method indicated in the row. The number of differentially expressed genes was identified using a t-test with an assumed significance level of $\alpha<0.05$. Q, quantile; AML, acute myeloid leukemia.

by real-time polymerase chain reaction (PCR) analysis (data not shown) or described in the literature as overexpressed (or, less frequently, underexpressed) in AML or immature hematopoietic cells, e.g., the CD34, enolase 1 (ENO1), azurocidin 1 (AZU1) and homeobox (HOX) genes. These genes were our strong candidates for differentially expressed genes. The negative controls constituted the probes that corresponded to the housekeeping genes e.g., glyceraldehyde-3-phosphate dehydrogenase, spermatogenic (GAPDHS), vimentin (VIM) and genes representing the phosphofructokinase (PFK) and ribosomal protein (RP) families. These housekeeping genes were expected to be expressed in both leukemic and healthy cells, but their expression levels should not differ between these two types of samples. In total, we selected 40 probes that served as the positive controls (Table VI) and 40 probes that served as the negative controls (Table VII). We then calculated the sensitivity and specificity of the normalization methods based on the percentage of positive controls that were present and the percentage of negative controls that were absent in each list of differentially expressed genes.

The graphical presentation shown in Fig. 3, as well as the data presented in Tables V, VI and VII, revealed the general relationship between the global loess normalization methods. The Snm2 and Spike methods yielded the most divergent results.

Table VI. The selected subset of 40 positive control probes used in the DEA of the AML dataset, which was normalized using ten double-channel and three single-channel methods.

Positive control probes	Double-channel normalization										Single-channel normalization		
	Spike	LoessM	Loess	Ploess	Vsn2	Nn	Olin	Olin_c	Turbo	Snm2	Vsn1	Q	Snm1
AZU1	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
BCL2	No	No	No	No	No	Yes	No	No	No	No	No	No	No
BCL2L1	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No
BTG1	No	No	No	No	Yes	Yes	No	No	Yes	No	Yes	Yes	Yes
CD34	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No
CD34_O	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No	No	No	No
CDK6	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No
CRYAA	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No	No	No	No
ENO1	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
FTL3	No	No	No	No	No	No	No	No	No	No	No	No	No
FLT3_O	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	No
GATA2	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No
GJB1	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes
HCK	No	Yes	No	No	No	Yes	Yes	Yes	Yes	No	No	No	No
HOXA10	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No
HOXA4	No	Yes	Yes	No	No	No	Yes	No	Yes	No	No	No	No
HOXA9	No	Yes	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
HOXB2	Yes	No	No	No	No	No	No	No	No	No	No	No	No
HOXB5	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
HOXB6	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
HRAS	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	Yes	Yes
JUNB	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No	No	No
KIT	No	Yes	No	No	Yes	No	Yes	Yes	No	No	Yes	Yes	No
LTB	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	No
MLLT1_O	No	Yes	Yes	Yes	Yes	No	No	No	Yes	No	Yes	No	Yes
MLLT10	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes
MLLT4	No	No	No	No	No	Yes	No	No	No	No	Yes	Yes	No
MN1	No	Yes	Yes	No	Yes	No	No	No	Yes	No	No	No	Yes
MPO	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	No	Yes
NPM1	No	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	No	No	No	No
PDE3B	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No	No	No
PF4	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No
PIM1	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	Yes	Yes
PRG1	No	Yes	Yes	No	Yes	No	No	Yes	Yes	No	Yes	Yes	Yes
S1000A8_O	No	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	No	**Yes**	**Yes**	No
S1000A9	No	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	No	No	No	No
S1000A9_O	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	No	**Yes**	**Yes**	No
SET	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
STMN1	No	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	No	**Yes**	**Yes**	No
TUBB	No	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes
No. of positive controls	4	35	31	25	30	27	30	32	34	1	19	19	16
Sensitivity (%)	10	87.5	77.5	62.5	75	67.5	75	80	85	2.5	47.5	47.5	40
Rank	9	1	4	8	5	7	5	3	2	10	1	1	3

The probe names are arranged in alphabetical order. The five probes indicated by bold print correspond to the four genes that were validated by real-time PCR analysis. All four genes showed differences in their expression level between the leukemic and control cells, but only the differences obtained with STMN1 and S100A9 were statistically significant. DEA, differential expression analysis; Q, quantile; AML, acute myeloid leukemia; PCR, polymerase chain reaction.

Table VII. The selected subset of negative control probes used in the DEA of the AML dataset, which was normalized using ten double-channel and three single-channel methods.

Negative control probes	Double-channel normalization										Single-channel normalization		
	Spike	LoessM	Loess	Ploess	Vsn2	Nn	Olin	Olin_c	Turbo	Snm2	Vsn1	Q	Snm1
AAMP	No	No	No	No	No	No	No	No	No	No	No	No	No
ACTG1	Yes	No	No	No	No	Yes	No	No	No	No	No	No	No
ALDOC	No	No	No	No	No	Yes	No	No	No	No	Yes	Yes	No
ARF1	No	No	No	No	No	No	No	No	No	No	No	No	No
CANX	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No
CLU	No	No	No	No	No	No	No	No	No	No	No	No	No
FTL	No	No	No	No	No	No	No	No	No	No	Yes	Yes	Yes
G6PD	No	No	No	No	No	No	No	Yes	No	No	No	No	No
GAPDHS	No	No	No	No	No	No	No	No	No	No	No	No	No
H3F3A	No	No	No	No	No	No	No	No	No	No	No	No	No
H3F3A_O	No	Yes	Yes	Yes	No	Yes	No	Yes	No	No	No	No	No
HPRT1	No	No	No	No	No	No	No	No	No	No	No	No	No
HSP90AA1	No	No	No	No	No	No	No	No	No	No	No	No	No
LDHA	No	No	No	No	No	No	No	No	No	Yes	No	No	No
LDHALGA	No	No	No	No	No	No	No	No	No	No	No	No	No
LDHC	No	No	No	No	No	No	No	No	No	No	No	No	No
MONO	No	No	No	No	No	No	No	No	No	No	No	No	No
MT2A	No	No	No	No	No	No	No	No	No	No	Yes	Yes	No
NONO_O	Yes	No	No	No	No	No	No	No	No	No	No	No	No
PFKL	No	No	No	No	No	Yes	No	No	No	No	No	No	No
PFKM	No	No	No	No	No	No	No	No	No	No	No	No	No
PFKP	No	No	No	No	No	No	No	No	No	No	No	No	No
PGAM1	No	No	No	No	No	No	No	No	No	No	No	No	No
PGK1	No	Yes	No	No	No	No	No	No	Yes	No	No	No	Yes
PGK2	No	No	No	No	No	No	No	No	No	No	No	No	No
RAC2	No	No	No	No	No	No	No	No	No	No	No	No	No
RPL0	No	No	No	No	No	No	No	No	No	No	No	No	No
RPL11	No	No	No	No	No	No	No	No	No	No	No	No	No
RPL19	No	No	No	No	No	No	No	No	No	No	No	No	No
RPL37A	No	No	No	Yes	No	No	No	No	No	Yes	Yes	Yes	No
RPL5	No	No	No	No	No	No	No	No	No	No	No	No	No
RPLP1_	No	No	No	No	No	No	No	No	No	No	Yes	No	Yes
RPS27A	No	No	No	No	No	No	No	No	No	No	No	No	No
RPS29	No	No	No	No	No	No	No	No	No	No	No	No	No
RPS3	No	No	No	No	No	Yes	No	No	No	No	No	No	No
TCEA1	No	No	No	No	No	No	No	No	No	No	No	No	Yes
TCFL1	No	No	No	No	No	No	No	No	No	No	No	No	No
TMSB4X	No	No	No	No	No	No	No	No	No	No	No	No	No
TUBA1	No	No	No	No	No	No	No	No	No	No	No	No	No
TUBB	No	No	No	No	No	No	No	No	No	No	No	No	No
No. of absent negative controls	38	37	38	37	39	34	39	37	39	38	35	36	36
Specificity (%)	95	92.5	95	92.5	97.5	85	97.5	92.5	97.5	95	87.5	90	90
Rank	4	7	4	7	1	10	1	7	1	4	3	1	1

The probe names are arranged in alphabetical order. DEA, differential expression analysis; Q, quantile; AML, acute myeloid leukemia; PCR, polymerase chain reaction.

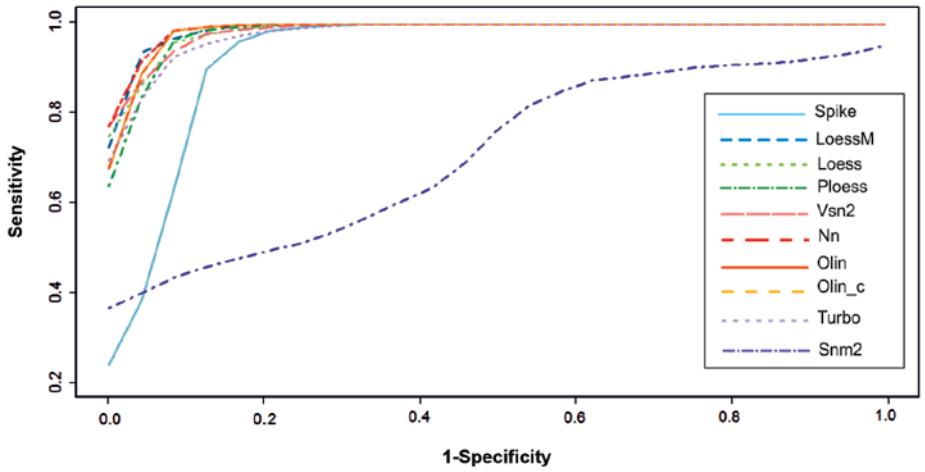

Figure 4. Receiver operating characteristic (ROC) curves for the 35-gene-based classifiers identified in the acute myeloid leukemia (AML) dataset following normalization using the double-channel normalization methods.

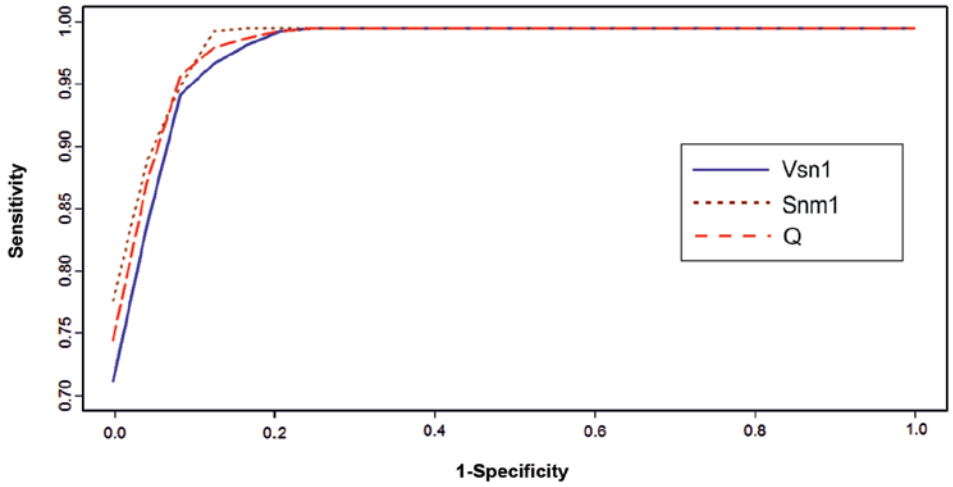

Figure 5. Receiver operating characteristic (ROC) curves for the 35-gene-based classifiers identified in the acute myeloid leukemia (AML) dataset following normalization using the single-channel normalization methods.

In our analysis, the most efficient method was Turbo, followed by LoessM; however, all of the double-channel normalization methods, with the exception of Snm2 and Spike, worked well. At least two-thirds of the positive controls and not more than 15% of the negative controls were selected from data obtained after the normalization of the AML dataset using these methods. The single-channel methods generated slightly worse results, but it should be noted that the set of 36 genes that were commonly found by ten of the normalization methods (eight double-channel and two single-channel methods) (Fig. 3) contains three probes that are complementary to the genes confirmed as differentially expressed by real-time PCR analysis.

ROC curves and AUC values. ROC curves are often used to test the power of classifiers, e.g., based on gene expression signatures. However, a comparison of ROC curves is only possible if these curves are generated for the same number of parameters. As shown in Table V, the lowest number of differentially expressed genes is 35 (obtained using the Snm2 method). To be able to compare all the normalization methods using ROC curves, we used 35 genes from the top of the lists obtained after each normalization method to generate the respective ROC curves. The ROC curves for the double- and single-channel normalization methods were generated independently (Figs. 4 and 5).

The results presented in Fig. 4 clearly show that the ROC curves for all of the methods (except Snm2 and Spike) have very similar shapes and occasionally overlap each other. A similar trend was observed with the three single-channel methods. To differentiate among the similar ROC curves, we calculated the AUC values. The results are presented in Table VIII.

Based on the results presented in Table VIII, we concluded that the Ploess, Olin, Olin_c and Nn methods exhibited the best performance relative to the other double-channel normalization methods (AUC values = 1). However, all the methods, with the exception of Snm2, had considerably high AUC values (>0.9), and seven methods had AUC values ≥0.99. All the tested single-channel normalizations were characterized by very high AUC values.

The final ranking. Combining all of the criteria described in the previous sections, we were able to determine which of the

Table VIII. AUC values (rounded to two decimal places) obtained after the normalization of the AML dataset using each normalization method.

	Double-channel normalization methods										Single-channel normalization methods		
	Spike	LoessM	Loess	Ploess	Vsn2	Nn	Olin	Olin_c	Turbo	Snm2	Vsn1	Q	Snm1
AUC value	0.91	0.99	0.99	1	0.99	1	1	1	0.98	0.58	0.99	1	1
Rank	9	5	5	1	5	1	1	1	8	10	3	1	1

These values were calculated for the classifiers consisting of the top 35 differentially expressed genes, which were selected using a t-test. AUC, area under the curve; Q, quantile; AML, acute myeloid leukemia.

Table IX. The final rank of the normalization methods used for the normalization of the AML dataset.

	Rank						
			DEA				
Normalization methods	Bias	Variance	Sensitivity	Specificity	AUC	Mean	Final rank
Double-channel normalization methods							
Spike	10	10	9	4	9	8.4	10
LoessM	8	7	1	7	5	5.6	6
Loess	2	2	4	4	5	3.4	3
Ploess	7	9	8	7	1	6.4	8
Vsn2	4	3	5	1	5	3.6	4
Nn	6	6	7	10	1	6	7
Olin	3	4	5	1	1	2.8	2
Olin_c	5	5	3	7	1	4.2	5
Turbo	1	1	2	1	8	2.6	1
Snm2	9	8	10	4	10	8.2	9
Single-channel normalization methods							
Vsn1	2	2	1	3	3	2.2	2
Q	1	1	1	1	1	1	1
Snm1	3	3	3	1	1	2.2	2

The final rank is based on the bias and variance values, the DEA and the AUC values. DEA, differential expression analysis; AUC, area under the curve; Q, quantile; AML, acute myeloid leukemia.

tested methods would be appropriate for the normalization of the AML dataset. Table IX summarizes the ranks obtained based on the bias and variance values, the results of the DEA and the AUC values (the data were obtained from Tables III, IV, VI, VII and VIII, respectively). The final rank was based on the mean of the ranks established separately for each criterion.

According to the results collected in Table IX for the double-channel methods, the global methods based on a local regression had the highest ranks. Turbo and Olin had the highest ranking, followed by the Loess, Vsn2 and the second method in the olin package (Olin_c). The Spike and Snm2 normalization methods occupied the last two places in the ranking. In the case of the single-channel normalization methods, the best results were obtained with the Q method; however, it must be noted that the differences between the Q method and the two other single-channel normalization methods, particularly Vsn1, were subtle. It is possible that the normalization of other boutique array datasets can yield results that are different from those obtained with the AML dataset. Therefore, we suggest the application of the following workflow to determine which normalization method is optimal for a specific dataset: i) Normalize the data using a few candidate methods, ii) calculate the 'bias' in the normalized data based on equations [1] or [5]. Rank the methods based on the bias values, iii) calculate the 'variance' in the normalized data based on equations [2] or [6]. Rank the methods based on the variance values, iv) identify

the differentially expressed genes after the dataset is normalized. Make a list of the differentially expressed genes with each normalization method, v) select a subset of probes that can serve as positive and negative controls to investigate the sensitivity and specificity of normalization methods. Rank the methods based on the sensitivity and specificity, vi) draw ROC curves and calculate the AUC values based on the number of differentially expressed genes that was selected by the most restrictive normalization algorithm. Rank the methods based on the AUC values, vii) for each normalization method, combine the ranks obtained using all the calculated criteria to determine the final rank. The method with the highest rank in the final ranking is considered to be the most appropriate method for the normalization of the investigated dataset.

The diagnostic plots could serve as additional determinants and may be helpful for the rejection of the most outstanding methods that evidently fail.

Discussion

The power of microarray technology resides in its complexity, availability, wide range of applications, miniaturization and relatively low cost compared with other technologies, e.g., deep sequencing. Small 'boutique' arrays are less expensive than whole-genome arrays and can be more easily applied for diagnostic or prognostic purposes. However, these arrays require more careful data pre-processing. All microarray data must be normalized prior to further analysis. The danger with normalization lies in the reduction of variability, such that technical bias and biological differences are no longer visible. This risk is of particular concern in the analysis of small, focused microarrays.

Although several methods have been designed or modified to address the issues encountered in the normalization of the data generated by boutique DNA microarrays, not all of these methods can be used for all datasets, e.g., the wlowess method proposed by Oshlack *et al* (47). The wlowess normalization procedure is based on the implementation of quantitative weights in an intensity-dependent normalization and is an alternative to the robust composite method elaborated by Yang *et al* (39). However, both the wlowess and composite strategies require a special set of control probes, i.e., a titration series of a whole microarray transcript pool, which are in principle dedicated to cDNA arrays (47). More often, such as in our three experiments, control oligonucleotide probes complementary to housekeeping genes and spike RNAs are used.

Usually, the selection of a normalization method is supported by the analysis of diagnostic diagrams, such as box plots, scatter plots or modified scatter plots (MAplots). However, the use of graphical tools is intuitive and often based on the subjective perception of the researcher. In our opinion, a much more elegant and objective solution was proposed in the study by Argyropoulos *et al* (67), who indicated that the most important aspects of the selection of a normalization algorithm are accuracy, precision and over-fitting. These three aspects can be verified using the bias, variance and relative entropy, respectively, which can be calculated by mathematical formulae. The bias and variance values are calculated for the control probes. A lower bias indicates greater accuracy, a lower variance indicates greater precision, and a lower relative entropy of the log-ratio distribution indicates a lower over-normalization potential. The estimation of the over-normalization by the relative entropy measure is possible only when self-self hybridization data are available.

Argyropoulos *et al* (67) compared three normalization strategies, i.e., global median, spike-based and loess, and found that global median normalization and spike-based normalization yield similar results and may be used interchangeably when the differentially expressed genes are identified based on statistical methods rather than intensity thresholds. With respect to the bias, the superiority of the loess method was clear. However, the loess method also exhibited the highest contribution to over-fitting, which drastically reduced the initial variability in the data. Taking into account all the above parameters, the spike-based normalization method was selected as the most appropriate for the tested dataset. However, Argyropoulos *et al* (67) admitted that there is no universally optimal normalization strategy.

In this study, we aimed to test the normalization methods that are available in R Bioconductor, which is a free, open-source software that offers a wide range of tools for microarray data analysis (http://www.bioconductor.org). Based on various criteria (bias and variance values for control probes, sensitivity and specificity of normalization methods and ROC curve analysis), we presented the results for ten double-channel and three single-channel normalization methods in the analysis of four different microarray datasets. Some of the tested algorithms, e.g., Loess and Vsn, are quite well known and popular, whereas others, e.g., Turbo, Olin, Nn and Snm, are rarely used. Comparing the bias and variance values enabled us to pre-select the best-performing methods, i.e., Turbo and Loess, for the analysis of two of our datasets (the AML and Allergy datasets), Ploess for analysis of the Asthma dataset and Ploess and Loess for analysis of the Oshlack dataset. We showed that simultaneous monitoring of diagnostic plots was helpful but not necessary because the conclusions drawn from the inspection of the plots and from the bias and variance calculations were consistent.

Using the AML dataset, which is devoted to gene expression studies of AML, we showed the impact of each normalization procedure on the differential analysis results. The application of a criterion based on biological relevance demands some background knowledge but can also be supported by ROC curve analysis. Calculating the AUC values is another approach for the mathematical and unsupervised classification of the tested methods. The combination of all of the parameters revealed that the global loess methods, such as Turbo from the TurboNorm package, Olin from the OLIN package and Loess from the limma package, were the most appropriate for the normalization of the AML two-color data. The lower efficiency of the Ploess method for the normalization of the AML dataset relative to other normalization methods based on locally weighted regression can be easily explained. According to Smyth (61), Ploess is not appropriate for microarrays that do not have print-tip groups (e.g., arrays in which the probes are synthesized *in situ*) or for small density-spotted arrays with a low number of spots per print-tip (<150). The AML dataset contained only 81 spots per print-tip. In the case of the Allergy and Asthma datasets, one print-tip counted 144 spots, which seemed to be sufficient to efficiently normalize one, but not

both, of these datasets. In the case of the dataset described in the study by Oshlack *et al* (47), in which the number of spots per print-tip was equal to 462, Ploess was the first choice in terms of bias and variance.

A somewhat surprising result was obtained for the Spike method, which is superficially the most sensible normalization strategy for boutique arrays. The failure of this method in the normalization of the AML, Asthma and Allergy datasets was likely the result of an insufficient number of spike-in control probe replicates across the arrays. However, in the case of the microarrays from the dataset from the study by Oshlack *et al* (47), which do not suffer from this problem, the Spike method was also ranked last with regard to bias and variance. Consequently, it can be concluded that normalization based on control probes requires special care.

The application of single-channel normalization approaches to double-channel microarray data can sometimes be a reasonable solution, e.g., when the variance in one channel (usually the reference) is much higher than that in the second channel or when the data from dye-swapped microarrays are not provided. In such cases, the differences in the dye decomposition or in the labeling efficiency cannot be eliminated in the normalization step. In contrast, ignoring the data collected from one channel could result in overlooking other sources of systematic variation, which can be introduced during microarray printing or hybridization. The opposite strategies are also known, i.e., the application of methods designed for two-color data, such as loess smoothing, in the pre-processing of one-color cDNA microarray data (71). The selection of the normalization option (single- or double-channel) when both are available remains the decision of the researcher. Of the single-channel normalization methods, Q from the limma package was the most optimal for the normalization of the AML dataset.

In conclusion, we demonstrate that boutique arrays, even atypical arrays, can be successfully normalized using standard methods that were initially developed for the analysis of oligonucleotide or cDNA microarrays covered with a high number of probes to represent the whole genome/transcriptome. However, not all of the available methods are suitable for each dataset. To obtain the most reliable results, it is necessary to carefully consider the first stages of data processing. The present study may prove useful for the selection of the optimal normalization strategy for any dataset.

Acknowledgements

The present study was supported by grant No. PBZ-MNiI-2/1/2005, from the Polish Ministry of Science and Informatization in 2006-2010 to M.F., and grant No. N407 061 32/2710, from the Polish Ministry of Science and Higher Education in 2007-2010 to M.M.S., and grant No. 2011/01/B/NZ5/02773, which was funded by the National Science Centre to P.K. We would like to thank Dr Beata Cudowska and Professor Maciej Kaczmarski from the Medical University of Bialystok, Białystok, Poland for providing the samples included in the Allergy dataset and Dr Aleksandra Szczepankiewicz, Dr Ewa Schöneich and Professor Anna Bręborowicz from Poznan University of Medical Sciences, Poznan, Poland for providing the samples included in the Asthma dataset. We are also grateful to Dr Marcin Schmidt from Poznan University of Life Sciences, Poznan, Poland for the HL60 cell culture preparation. B.U. is a scholarship holder of 'Scholarship for PhD candidates in the fields considered strategic for the development of Greater Poland', project edition 2011/2012.

References

1. Howbrook DN, van der Valk AM, O'Shaugnessy MC, Sarker DK, Baker SC and Lloyd AW: Developments in microarray technologies. Drug Discov Today 8: 642-651, 2003.
2. Venkatasubbarao S: Microarrays - status and prospects. Trends Biotechnol 22: 630-637, 2004.
3. Meloni R, Khalfallah O and Biguet NF: DNA microarrays and pharmacogenomics. Pharmacol Res 49: 303-338, 2004.
4. Trevino V, Falciani F and Barrera-Saldaña HA: DNA microarrays: a powerful genomic tool for biomedical and clinical research. Mol Med 13: 527-541, 2007.
5. Malone JH and Oliver B: Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biol 9: 34, 2011.
6. Mao X, Young BD and Lu YJ: The application of single nucleotide polymorphism microarrays in cancer research. Curr Genomics 8: 219-228, 2007.
7. Shinawi M and Cheung SW: The array CGH and its clinical applications. Drug Discov Today 13: 760-770, 2008.
8. Han X, Lin X, Liu B, Hou Y, Huang J, Wu S, Liu J, Mei L, Jia G and Zhu Q: Simultaneously subtyping of all influenza A viruses using DNA microarrays. J Virol Methods 152: 117-121, 2008.
9. Keren B and Le Caignec C: Oligonucleotide microarrays in constitutional genetic diagnosis. Expert Rev Mol Diagn 11: 521-532, 2011.
10. Chang HY, Sneddon JB, Alizadeh AA, Sood R, West RB, Montgomery K, Chi JT, van de Rijn M, Botstein D and Brown PO: Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. PLoS Biol 2: E7, 2004.
11. Nelson AM, Zhao W, Gilliland KL, Zaenglein AL, Liu W and Thiboutot DM: Temporal changes in gene expression in the skin of patients treated with isotretinoin provide insight into its mechanism of action. Dermatoendocrinol 1: 177-187, 2009.
12. Verma G, Bhatia H and Datta M: Gene expression profiling and pathway analysis identify the integrin signaling pathway to be altered by IL-1β in human pancreatic cancer cells: role of JNK. Cancer Lett 320: 86-95, 2012.
13. Lu T, Costello CM, Croucher PJ, Häsler R, Deuschl G and Schreiber S: Can Zipf's law be adapted to normalize microarrays? BMC Bioinformatics 6: 37, 2005.
14. Mi S, Lu J, Sun M, Li Z, Zhang H, Neilly MB, Wang Y, Qian Z, Jin J, Zhang Y, Bohlander SK, Le Beau MM, Larson RA, Golub TR, Rowley JD and Chen J: MicroRNA expression signatures accurately discriminate acute lymphoblastic leukemia from acute myeloid leukemia. Proc Natl Acad Sci USA 104: 19971-19976, 2007.
15. Garzon R, Volinia S, Liu CG, Fernandez-Cymering C, Palumbo T, Pichiorri F, Fabbri M, Coombes K, Alder H, Nakamura T, Flomenberg N, Marcucci G, Calin GA, Kornblau SM, Kantarjian H, Bloomfield CD, Andreeff M and Croce CM: MicroRNA signatures associated with cytogenetics and prognosis in acute myeloid leukemia. Blood 111: 3183-3189, 2008.
16. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD and Lander ES: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286: 531-537, 1999.
17. Borczuk AC, Shah L, Pearson GD, Walter KL, Wang L, Austin JH, Friedman RA and Powell CA: Molecular signatures in biopsy specimens of lung cancer. Am J Respir Crit Care Med 170: 167-174, 2004.
18. Lapointe J, Li C, Higgins JP, van de Rijn M, Bair E, Montgomery K, Ferrari M, Egevad L, Rayford W, Bergerheim U, Ekman P, DeMarzo AM, Tibshirani R, Botstein D, Brown PO, Brooks JD and Pollack JR: Gene expression profiling identifies clinically relevant subtypes of prostate cancer. Proc Natl Acad Sci USA 101: 811-816, 2004.
19. Haferlach T, Kohlmann A, Schnittger S, Dugas M, Hiddemann W, Kern W and Schoch C: Global approach to the diagnosis of leukemia using gene expression profiling. Blood 106: 1189-1198, 2005.

20. Dudaladava V, Jarzab M, Stobiecka E, Chmielik E, Simek K, Huzarski T, Lubiński J, Pamuła J, Pekala W, Grzybowska E and Lisowska K: Gene expression profiling in hereditary, BRCA1-linked breast cancer: preliminary report. Hered Cancer Clin Pract 4: 28-38, 2006.
21. Park MH, Cho SA, Yoo KH, Yang MH, Ahn JY, Lee HS, Lee KE, Mun YC, Cho DH, Seong CM and Park JH: Gene expression profile related to prognosis of acute myeloid leukemia. Oncol Rep 18: 1395-1402, 2007.
22. Reis-Filho JS and Pusztai L: Gene expression profiling in breast cancer: classification, prognostication, and prediction. Lancet 378: 1812-1823, 2011.
23. Benes V and Muckenthaler M: Standardization of protocols in cDNA microarray analysis. Trends Biochem Sci 28: 244-249, 2003.
24. Kreil DP and Russel RR: There is no silver bullet - a guide to low-level data transforms and normalisation methods for microarray data. Brief Bioinform 6: 86-97, 2005.
25. Imbeaud S and Auffray C: 'The 39 steps' in gene expression profiling: critical issues and proposed best practices for microarray experiments. Drug Discover Today 10: 1175-1182, 2005.
26. Ness SA: Microarray analysis: basic strategies for successful experiments. Mol Biotechnol 36: 205-219, 2007.
27. Hartemink AJ, Gifford DK, Jaakkola TS and Young RA: Maximum likelihood estimation of optimal scaling factors for expression array normalization. SPIE BiOS. Proc SPIE 4266: 132-140, 2001.
28. Knapen D, Vergauwen L, Laukens K and Blust R: Best practices for hybridization design in two-colour microarray analysis. Trends Biotechnol 27: 406-414, 2009.
29. Margaritis T, Lijnzaad P, van Leenen D, Bouwmeester D, Kemmeren P, van Hooff SR and Holstege FC: Adaptable gene-specific dye bias correction for double-channel DNA microarrays. Mol Syst Biol 5: 266, 2009.
30. Dudoit S, Yang YH, Callow MJ and Speed TP: Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. Stat Sin 12: 111-139, 2002.
31. Bolstad BM, Irizarry RA, Astrand M and Speed TP: A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19: 185-193, 2003.
32. Smyth GK and Speed TP: Normalization of cDNA microarray data. Methods 31: 265-273, 2003.
33. Reimers M: Statistical analysis of microarray data. Addict Biol 10: 23-35, 2005.
34. Ritchie ME, Silver J, Oshlack A, Holmes M, Diyagama D, Holloway A and Smyth GK: A comparison of background correction methods for two-colour microarrays. Bioinformatics 23: 2700-2707, 2007.
35. Siatkowski I, Zyprych J, Handschuh L and Figlerowicz M: The methods of normalization used in the analysis of two-color microarrays. Colloquium Biometricum 39: 9-19, 2009.
36. Hahne F, Huber W, Gentleman R and Falcon S: Bioconductor Case Studies. Springer, New York, 2008.
37. Yang YH, Dudoit S, Luu P and Speed TP: Normalization for cDNA microarray data. In: Microarrays: Optical Technologies and Informatics. Bittner ML, Chen Y, Dorsel AN and Dougherty ER (eds). Proc SPIE 4266: pp141-152, 2001.
38. Quackenbush J: Microarray data normalization and transformation. Nat Genet 32 (Suppl): S496-S501, 2002.
39. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J and Speed TP: Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 30: e15, 2002.
40. Chua, S, Vijayakumar, P, Nissom P and Yang H: A novel normalization method for effective removal of systematic variation in microarray data. Nucleic Acids Res 34: e38, 2006.
41. Pelz CR, Kulesz-Martin M, Bagby G and Sears RC: Global rank-invariant set normalization (GRSN) to reduce systematic distortions in microarray data. BMC Bioinformatics 9: 520, 2008.
42. Cleveland WS: Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 74: 829- 836, 1979.
43. Cleveland WS: Lowess - a program for smoothing scatterplots by robust locally weighted regression. Am Stat 35: 54, 1981.
44. Millenaar FF, Okyere J, May ST, van Zanten M, Voesenek LA and Peeters AJ: How to decide? Different methods of calculating gene expression from short oligonucleotide array data will give different results. BMC Bioinformatics 7: 137, 2006.
45. Hsieh WP, Chu TM, Lin YM and Wolfinger RD: Kernel density weighted loess normalization improves the performance of detection within asymmetrical data. BMC Bioinformatics 12: 222, 2011.
46. Choe SE, Boutros M, Michelson AM, Church GM and Halfon MS: Preferred analysis methods for Affymetrix GeneChips revealed by a wholly defined control dataset. Genome Biol 6: R16, 2005.
47. Oshlack A, Emslie D, Corcoran LM and Smyth GK: Normalization of boutique two-color microarrays with a high proportion of differentially expressed probes. Genome Biol 8: R2, 2007.
48. Hoffman R, Seidl T and Dugas M: Profound effect of normalization on detection of differentially expressed genes in oligonucleotide microarray data analysis. Genome Biol 3: Research0033, 2002.
49. Schmidt MT, Handschuh L, Zyprych J, Szabelska A, Olejnik-Schmidt AK, Siatkowski I and Figlerowicz M: Impact of DNA microarray data transformation on gene expression analysis - comparison of two normalization methods. Acta Biochim Pol 58: 573-580, 2011.
50. Barash Y, Dehan E, Krupsky M, Franklin W, Geraci M, Friedman N and Kaminski N: Comparative analysis of algorithms for signal quantitation from oligonucleotide microarrays. Bioinformatics 20: 839-846, 2004.
51. Campanaro S, Romualdi C, Fanin M, Celegato B, Pacchioni B, Trevisan S, Laveder P, De Pittà C, Pegoraro E, Hayashi YK, Valle G, Angelini C and Lanfranchi G: Gene expression profiling in dysferlinopathies using a dedicated muscle microarray. Hum Mol Genet 11: 3283-3298, 2002.
52. McIlroy D, Tanguy-Royer S, Le Meur N, Guisle I, Royer PJ, Léger J, Meflah K and Grégoire M: Profiling dendritic cell maturation with dedicated microarrays. J Leukoc Biol 78: 794-803, 2005.
53. Ferrarini A, De Stefano M, Baudouin E, Pucciariello C, Polverari A, Puppo A and Delledonne M: Expression of Medicago truncatula genes responsive to nitric oxide in pathogenic and symbiotic conditions. Mol Plant Microbe Interact 21: 781-790, 2008.
54. Baron D, Magot A, Ramstein G, Steenman M, Fayet G, Chevalier C, Jourdon P, Houlgatte R, Savagner F and Pereon Y: Immune response and mitochondrial metabolism are commonly deregulated in DMD and aging skeletal muscle. PLoS One 11: e26952, 2011.
55. Held M, Gase K and Baldwin IT: Microarrays in ecological research: a case study of a cDNA microarray for plant-herbivore interactions. BMC Ecol 4: 13, 2004.
56. Wilson DL, Buckley MJ, Helliwell CA and Wilson IW: New normalization methods for cDNA microarray data. Bioinformatics 19: 1325-1332, 2003.
57. Wenne R, Handschuh L, Pocwierz-Kotus A, Urbaniak R, Formanowicz P, Całkiewicz J, Brzozowska K, Figlerowicz M, Węgrzyn G and Wróbel B: The application of microarray technology to the identification of Tc1-like element sequences in fish genomes. Marine Biology Res 7: 466-477, 2011.
58. Zmieńko A, Guzowska-Nowowiejska M, Urbaniak R, Pląder W, Formanowicz P and Figlerowicz M: A tiling microarray for global analysis of chloroplast genome expression in cucumber and other plants. Plant Methods 7: 29, 2011.
59. Gentleman R, Irizarry RA, Carey VJ, Dudoit S and Hubrer W: Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Springer, New York, 2005.
60. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY and Zhang J: Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5: R80, 2004.
61. Smyth GK: Limma: linear models for microarray data. In: Bioinformatics and Computational Biology Solutions using R and Bioconductor. Gentleman R, Carey RV, Dudoit S, Irizarry R and Huber W (eds). Springer, New York, 397-420, 2005.
62. Mecham BH, Nelson PS and Storey JD: Supervised normalization of microarrays. Bioinformatics 26: 1308-1315, 2010.
63. Huber W, von Heydebreck A, Sültmann H, Poustka A and Vingron M: Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 18 (Suppl 1): S96-S104, 2002.
64. Tarca AL, Cooke JE and Mackay J: A robust neural networks approach for spatial and intensity-dependent normalization of cDNA microarray data. Bioinformatics 21: 2674-2683, 2005.

65. Futschik ME and Crompton T: OLIN: optimized normalization, visualization and quality testing of two-channel microarray data. Bioinformatics 21: 1724-1726, 2004.
66. van Iterson M, Duijkers FA, Meijerink JP, Admiraal P, van Ommen GJ, Boer JM, van Noesel MM and Menezes RX: A novel and fast normalization method for high-density arrays. Stat Appl Genet Mol Biol 11: 1544-6115, 2012.
67. Argyropoulos C, Chatziioannou A, Nikifordis G, Moustakas A, Kollias G and Aidinis V: Operational criteria for selecting a cDNA microarray data normalization algorithm. Oncol Rep 15: 983-996, 2006.
68. Benjamini Y and Hochberg Y: Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc 57: 289-300, 1995.
69. Wolpert DH: Stacked generalization. Neural Networks 5: 241-259, 1992.
70. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC and Müller M: pROC: an open-source package for R and S^+ to analyze and compare ROC curves. BMC Bioinformatics 12: 77, 2011.
71. Edwards D: Non-linear normalization and background correction on single-channel cDNA microarray studies. Bioinformatics 19: 825-833, 2003.
72. Yang YH, Paquet A and Dudoit S: Exploratory analysis for two-color spotted microarray data. R package version 1.38.0.